

OHSU Informatics Analytics BD2K Skill Course

Oregon Health and Sciences University

PI: David A. Dorr, Melissa Haendel, and Shannon McWeeney

Grant Number: 1R25EB020379-01

We propose to create and disseminate an OHSU Informatics Analytics BD2K Skill Course to enable researchers and students at any career stage to gain crucial analytics and data skills and competencies as part of the Big Data to Knowledge (BD2K) initiative. Although the potential of big data to advance research is enormous, our needs assessment has identified that a wide range of researchers and graduate students lack fundamental skills in identifying the appropriate data sets to answer their hypotheses, in managing and 'wrangling' data to prepare it for analysis, and in using both common and advanced methods to analyze the data. These needs varied by experience and skill of the researcher. We therefore propose three educational components: the first is to recruit a diverse set of participants and work to inculcate a fundamental set of competencies in basic big data skills; the second is to create realistic but synthetic large datasets with embedded signals where students can practice their skills in collaborative and interactive exercises; the third is to create advanced challenges to help researchers test and implement more advanced methodologies. We anticipate the impact to be substantial in enabling researchers to improve their use of big data to more rapidly and accurately address important research hypotheses. We will disseminate this work via online curricula resources and public venues and share the datasets created so that others may train their own teams. PUBLIC HEALTH RELEVANCE: The importance of this project to public health is related to how big data will be used to advance health. There are tremendous opportunities to advance our understanding of factors influencing population health and treatment of illness that will come from the large datasets being generated; this work will enable current and future researchers to become skilled in the use of these datasets and the methods used to analyze them.